

RAW SEQUENCE LISTING ERROR REPORT

BIOTECHNOLOGY
SYSTEMS
BRANCH



The Biotechnology Systems Branch of the Scientific and Technical Information Center (STIC) detected errors when processing the following computer readable form:

Application Serial Number: 09/698,213

Source: O/PE

Date Processed by STIC: 11/13/2000

THE ATTACHED PRINTOUT EXPLAINS DETECTED ERRORS.

PLEASE FORWARD THIS INFORMATION TO THE APPLICANT BY EITHER:

- 1) INCLUDING A COPY OF THIS PRINTOUT IN YOUR NEXT COMMUNICATION TO THE APPLICANT, WITH A NOTICE TO COMPLY or,
- 2) TELEPHONING APPLICANT AND FAXING A COPY OF THIS PRINTOUT, WITH A NOTICE TO COMPLY

FOR CRF SUBMISSION QUESTIONS, PLEASE CONTACT MARK SPENCER, 703-308-4212.

FOR SEQUENCE RULES INTERPRETATION, PLEASE CONTACT ROBERT WAX, 703-308-4216.

PATENTIN 2.1 e-mail help: patin21help@uspto.gov or phone 703-306-4119 (R. Wax)

PATENTIN 3.0 e-mail help: patin30help@uspto.gov or phone 703-306-4119 (R. Wax)

TO REDUCE ERRORED SEQUENCE LISTINGS, PLEASE USE THE CHECKER VERSION 3.0 PROGRAM, ACCESSIBLE THROUGH THE U.S. PATENT AND TRADEMARK OFFICE WEBSITE. SEE BELOW:

Checker Version 3.0

The Checker Version 3.0 application is a state-of-the-art Windows based software program employing a logical and intuitive user-interface to check whether a sequence listing is in compliance with format and content rules. Checker Version 3.0 works for sequence listings generated for the original version of 37 CFR §§1.821 – 1.825 effective October 1, 1990 (old rules) and the revised version (new rules) effective July 1, 1998 as well as World Intellectual Property Organization (WIPO) Standard ST.25.

Checker Version 3.0 replaces the previous DOS-based version of Checker, and is Y2K-compliant. Checker allows public users to check sequence listings in Computer Readable form (CRF) before submitting them to the United States Patent and Trademark Office (USPTO). Use of Checker prior to filing the sequence listing is expected to result in fewer errored sequence listings, thus saving time and money.

Checker Version 3.0 can be down loaded from the USPTO website at the following address:

<http://www.uspto.gov/web/offices/pac/checker>

OIPE

RAW SEQUENCE LISTING
 PATENT APPLICATION: US/09/698,213

DATE: 11/13/2000
 TIME: 15:56:38

Input Set : A:\Sequence Listing BioInformatics.txt
 Output Set: N:\CRF3\11132000\I698213.raw

Does Not Comply
 Corrected Diskette Needed

2 <110> APPLICANT: McIninch, James
 4 <120> TITLE OF INVENTION: COMPUTATIONAL NUCLEIC ACID CODING AND FEATURE ANALYSIS
 6 <130> FILE REFERENCE: 04983.0220.00US00
 C--> 8 <140> CURRENT APPLICATION NUMBER: US/09/698,213
 C--> 8 <141> CURRENT FILING DATE: 2000-10-30
 8 <160> NUMBER OF SEQ ID NOS: 4
 10 <170> SOFTWARE: PatentIn version 3.0
 12 <210> SEQ ID NO: 1
 13 <211> LENGTH: 2165
 14 <212> TYPE: DNA
 15 <213> ORGANISM: Arabidopsis thaliana
 17 <220> FEATURE:
 18 <221> NAME/KEY: unsure
 19 <222> LOCATION: (1)...(2165)
 20 <223> OTHER INFORMATION: Unsure at all n locations
 22 <220> FEATURE:
 23 <223> OTHER INFORMATION: Ecotype Landsberg, genomic DNA
 25 <400> SEQUENCE: 1
 26 tactcaaaaa tatattccat gcttaattag gccggattcg cggtgacgat gcaccaagag 60
 28 cggtttttcc gagcattgta ggcgcctc gccacaccg tqtgatggtt gggatgggac 120
 30 aaaaggatgc ttatgttga gacgaggtc aatcaaaacg tggatcttg actctgaagt 180
 32 acccaattga gcatggaatt gtttaataatt gggatgacat ggagaagatt tggcatcaca 240
 34 cttttacaa tgagcttcgt gttgcccctg aagaacatcc ggtctcttg accgaagctc 300
 36 ctctcaatcc gaaagctaac cgtgagaaga tgactcagat catgtttgag acattcaata 360
 38 ctcctgatat gtatgttgc attcaagctg ttctctcact ctatgccagt gcccgtaacta 420
 40 ctggtcagta cattaactaca ttctttttat accgtttggt tgaataaaaa ttcggttttg 480
 42 ttcgattcga gtttgcctc attattttta tttgtttggt taggtattgt ttggactcc 540
 44 ggagatggtg tgagccacac ggtaccaatc tacgaggggt atgcacttcc acacgcaatc 600
 46 ctgcgtcttg atcttgacgg tcgtgacctc accgaccacc ttatgaaaat cctgacagag 660
 48 cgtggttact ctttcaccac aactgctgag cgtgagattg tttagagacat gaaggagaaq 720
 50 ctctcttaca ttgccttga ctttgaacaa gagctcgaga cttccaaaac aagctcatcc 780
 52 gttgagaaga gcttcgagct gccagacggt caagtgatca ccacggggc agagcgtttc 840
 54 cgatgcccctg aagttctggt tcagccatcg atgatcggaa tggaaaatcc gggaaattcat 900
 56 gaaactactt acaactcaat catgaaatgt gatgtggata tcaggaagga tctttatgga 960
 58 aacattgtgc ttatgtgtg caccacaatg ttcatgaggga ttggtgatag gatgagtaaa 1020
 60 gagatcaacg cgttggtccc aagcagtatg aacatcaaaq tggtygctcc accggaagg 1080
 62 aagtacagtg tctggatcgg tggctctatc ttggtctccc tcagtacttt ccagcaggta 1140
 64 aattacttac tatacttaat acataaagtc tattagtgat ttgatgtata aagtgttaca 1200
 66 aaaatgtgtt ccaaatttgc agatgtggat tgcgaaagcg gagtatgatg aatctggacc 1260
 68 gtcaatcgtc cacaggaagt gcttcctgac aaaagtcacc aagtaaaaca agagcggtaa 1320
 70 aaattttgat atcagttttt caccctgaag ccagttgcta taattactca caacttctct 1380
 72 atttgtgttc ttttattctt gtcctctggt gttcatttta atctcttttt tgcaacaaag 1440
 74 caacttaaaa aaacagagca gtcattaaac gaatgttatt attatatata tgtatacata 1500
 76 ttatgtatata cccattatct cattaaaaa ttatcatat aaggatagga ttctatacat 1560
 78 cgatatatit attttgttga cactattcag cacatgctta tgtcttatct tgttagtata 1620
 80 tgaacccaaa gacaaataat agatgctaca aattgttttc tttgaagcaa aaatttcaat 1680
 82 cttaaaaattg tttttttcca ggttacacaa aaaaacttg tagtttgtaa attttctata 1740

OK 2-4

RAW SEQUENCE LISTING
PATENT APPLICATION: US/09/698,213

DATE: 11/13/2000
TIME: 15:56:38

Input Set : A:\Sequence Listing BioInformatics.txt
Output Set: N:\CRF3\11132000\I698213.raw

84 cuatttttggg gatctcaaca agaacatgaa cttcaacttc tagtcatatg acgacctgag 1800
86 tctgcgcggc lgtgaatctc ttgtctgcag taaatgttta caagtgggtg gtaaatggg 1860
88 actgattcaa aagctttaag aaatctacac atttcgtqaa attatttagc aqacttgata 1920
90 ttaaaaaatct aggataaaat gactatccaa agacaaatag gactgtttca catgttcccc 1980
92 tgattcttgt agctcataac tcatcagcag ttaacttttc tacctcatac acgctcgcaa 2040
94 tncgttttga attatcagct ntaatttttc taattctttg gaaattatta gcagctcgat 2100
96 caaatggggc atgqcttctt cttctatctg caactcatct aaactttcca tqaqaaaca 2160
98 aagct 2165
101 <210> SEQ ID NO: 2
102 <211> LENGTH: 423
103 <212> TYPE: PRT
104 <213> ORGANISM: Unknown
106 <220> FEATURE:
107 <223> OTHER INFORMATION: Describes a predicted protein sequence
109 <220> FEATURE:
110 <221> NAME/KEY: site
111 <222> LOCATION: (1)...(423)
112 <223> OTHER INFORMATION: A stop codon is predicted at all XAA locations
114 <440> SEQUENCE: 2
W--> 116 Xaa Arg Phe Phe Arg Ala Leu Xaa Ala Val Leu Ala Thr Pro Val Xaa
117 5 10 15
119 Trp Leu Gly Trp Asp Lys Arg Met Leu Met Leu Glu Thr Arg Leu Asn
120 20 25 30
W--> 122 Gln Asn Val Val Ser Xaa Leu Xaa Ser Thr Gln Leu Ser Met Glu Leu
123 35 40 45
125 Leu Ile Ile Gly Met Thr Trp Arg Arg Phe Gly Ile Thr Leu Ser Thr
126 50 55 60
W--> 128 Met Ser Phe Val Leu Pro Leu Lys Asn Ile Arg Xaa Leu Thr Glu Ala
129 65 70 75 80
131 Pro Leu Asn Pro Lys Ala Asn Arg Glu Lys Met Thr Gln Ile Met Phe
132 85 90 95
134 Glu Thr Phe Asn Thr Pro Ala Met Tyr Val Ala Ile Gln Ala Val Leu
135 100 105 110
137 Ser Leu Tyr Ala Ser Gly Arg Thr Thr Gly Gln Tyr Ile Thr Thr Phe
138 115 120 125
W--> 140 Phe Leu Tyr Arg Xaa Ser Gly Asp Gly Val Ser His Thr Val Pro Ile
141 130 135 140
143 Tyr Glu Gly Tyr Ala Leu Pro His Ala Ile Leu Arg Leu Asp Leu Ala
144 145 150 155 160
146 Gly Arg Asp Leu Thr Asp His Leu Met Lys Ile Leu Thr Glu Arg Gly
147 165 170 175
149 Tyr Ser Phe Thr Thr Ala Glu Arg Glu Ile Val Arg Asp Met Lys
150 180 185 190
152 Glu Lys Leu Ser Tyr Ile Ala Leu Asp Phe Glu Gln Glu Leu Glu Thr
153 195 200 205
155 Ser Lys Thr Ser Ser Ser Val Glu Lys Ser Phe Glu Leu Pro Asp Gly
156 210 215 220
158 Gln Val Ile Thr Ile Gly Ala Glu Arg Phe Arg Cys Pro Glu Val Leu
159 225 230 235 240

Xaa can only represent
a single amino acid;
it cannot represent
a stop codon,
which is not an
amino acid. FYI,
Per 1.822 of
Sequence Rules,

"An amino acid sequence
that contains internal
terminator symbols...
may not be represented
as a single amino acid
sequence, but shall be
presented as separate
amino acid sequences."

RAW SEQUENCE LISTING
 PATENT APPLICATION: US/09/698,213
 DATE: 11/13/2000
 TIME: 15:56:38

Input Set : A:\Sequence Listing BioInformatics.txt
 Output Set: N:\CRF3\11132000\I698213.raw

```

161 Phe Gln Pro Ser Met Ile Gly Met Glu Asn Pro Gly Ile His Glu Thr
162                245                250                255
164 Thr Tyr Asn Ser Ile Met Lys Cys Asp Val Asp Ile Arg Lys Asp Leu
165                260                265                270
167 Tyr Gly Asn Ile Val Leu Ser Gly Gly Thr Thr Met Phe Asp Gly Ile
168                275                280                285
170 Gly Asp Arg Met Ser Lys Glu Ile Thr Ala Leu Ala Pro Ser Ser Met
171                290                295                300
173 Lys Ile Lys Val Val Ala Pro Pro Glu Arg Lys Tyr Ser Val Trp Ile
174 305                310                315                320
W--> 176 Gly Gly Ser Ile Xaa Val Pro Asn Leu Gln Met Trp Ile Ala Lys Ala
177                325                330                335
W--> 179 Glu Tyr Xaa Asn Leu Asp Arg Gln Ser Ser Thr Gly Ser Ala Ser Asp
180                340                345                350
W--> 182 Gln Lys Ser Pro Ser Lys Thr Arg Ala Val Lys Ile Leu Xaa Asn Ser
183                355                360                365
185 Ser Ala Val Asn Phe Ser Thr Ser Tyr Thr Leu Ala Ile Arg Leu Glu
186                370                375                380
188 Leu Ser Ala Leu Ile Phe Leu Ile Ser Leu Glu Ile Ile Ser Ser Ser
189 385                390                395                400
191 Ile Lys Trp Gly Met Ala Ser Ser Ser Ile Cys Asn Ser Ser Lys Leu
192                405                410                415
W--> 194 Ser Met Lys Lys Gln Ser Xaa
195                420
197 <210> SEQ ID NO: 3
198 <211> LENGTH: 422
199 <212> TYPE: PRT
200 <213> ORGANISM: Unknown
202 <220> FEATURE:
203 <223> OTHER INFORMATION: Describes a predicted protein sequence
205 <220> FEATURE:
206 <221> NAME/KEY: site
207 <222> LOCATION: (1)...(422)
208 <223> OTHER INFORMATION: A stop codon is predicted at all XAA locations
211 <400> SEQUENCE: 3
W--> 213 Xaa Arg Phe Phe Arg Ala Leu Xaa Ala Val Leu Ala Thr Pro Val Xaa
214 1                5                10                15
216 Trp Leu Gly Trp Asp Lys Arg Met Leu Met Leu Glu Thr Arg Leu Asn
217                20                25                30
W--> 219 Gln Asn Val Val Ser Xaa Leu Xaa Ser Thr Gln Leu Ser Met Glu Leu
220                35                40                45
222 Leu Ile Ile Gly Met Thr Trp Arg Arg Phe Gly Ile Thr Leu Ser Thr
223                50                55                60
W--> 225 Met Ser Phe Val Leu Pro Leu Lys Asn Ile Arg Xaa Leu Thr Glu Ala
226 65                70                75                80
228 Pro Leu Asn Pro Lys Ala Asn Arg Glu Lys Met Thr Gln Ile Met Phe
229                85                90                95
231 Glu Thr Phe Asn Thr Pro Ala Met Tyr Val Ala Ile Gln Ala Val Leu
232                100                105                110

```

same error

RAW SEQUENCE LISTING

DATE: 11/13/2000

PATENT APPLICATION: US/09/698,213

TIME: 15:56:38

Input Set : A:\Sequence Listing BioInformatics.txt

Output Set: N:\CRF3\11132000\I698213.raw

234 Ser Leu Tyr Ala Ser Gly Arg Thr Thr Gly Gln Tyr Ile Thr Thr Phe
 235 115 120 125
 W--> 237 Phe Leu Tyr Arg Xaa Ser Gly Asp Gly Val Ser His Thr Val Pro Ile
 238 130 135 140
 240 Tyr Glu Gly Tyr Ala Leu Pro His Ala Ile Leu Arg Leu Asp Leu Ala
 241 145 150 155 160
 243 Gly Arg Asp Leu Thr Asp His Leu Met Lys Ile Leu Thr Glu Arg Gly
 244 165 170 175
 246 Tyr Ser Phe Thr Thr Thr Ala Glu Arg Glu Ile Val Arg Asp Met Lys
 247 180 185 190
 249 Glu Lys Leu Ser Tyr Ile Ala Leu Asp Phe Glu Gln Glu Leu Glu Thr
 250 195 200 205
 252 Ser Lys Thr Ser Ser Ser Val Glu Lys Ser Phe Glu Leu Pro Asp Gly
 253 210 215 220
 255 Gln Val Ile Thr Ile Gly Ala Glu Arg Phe Arg Cys Pro Glu Val Leu
 256 225 230 235 240
 258 Phe Gln Pro Ser Met Ile Gly Met Glu Asn Pro Gly Ile His Glu Thr
 259 245 250 255
 261 Thr Tyr Asn Ser Ile Met Lys Cys Asp Val Asp Ile Arg Lys Asp Leu
 262 260 265 270
 264 Tyr Gly Asn Ile Val Leu Ser Gly Gly Thr Thr Met Phe Asp Gly Ile
 265 275 280 285
 267 Gly Asp Arg Met Ser Lys Glu Ile Thr Ala Leu Ala Pro Ser Ser Met
 268 290 295 300
 270 Lys Ile Lys Val Val Ala Pro Pro Glu Arg Lys Tyr Ser Val Trp Ile
 271 305 310 315 320
 W--> 273 Gly Gly Ser Ile Leu Ala Ser Xaa Gln Met Trp Ile Ala Lys Ala Glu
 274 325 330 335
 W--> 276 Tyr Xaa Asn Leu Asp Arg Gln Ser Ser Thr Gly Ser Ala Ser Asp Gln
 277 340 345 350
 W--> 279 Lys Ser Pro Ser Lys Thr Arg Ala Val Lys Ile Leu Xaa Asn Ser Ser
 280 355 360 365
 282 Ala Val Asn Phe Ser Thr Ser Tyr Thr Leu Ala Ile Arg Leu Glu Leu
 283 370 375 380
 285 Ser Ala Leu Ile Phe Leu Ile Ser Leu Glu Ile Ile Ser Ser Ser Ile
 286 385 390 395 400
 288 Lys Trp Gly Met Ala Ser Ser Ser Ile Cys Asn Ser Ser Lys Leu Ser
 289 405 410 415
 W--> 291 Met Lys Lys Gln Ser Xaa
 292 420
 294 <210> SEQ ID NO: 4
 295 <211> LENGTH: 296
 296 <212> TYPE: PRT
 297 <213> ORGANISM: Arabidopsis thaliana
 299 <220> FEATURE:
 300 <223> OTHER INFORMATION: Ecotype columbia, describes actin
 302 <400> SEQUENCE: 4
 304 Met Glu Lys Ile Trp His His Thr Phe Tyr Asn Glu Leu Arg Val Ala
 305 1 5 10 15

same

RAW SEQUENCE LISTING

DATE: 11/13/2000

PATENT APPLICATION: US/09/698,213

TIME: 15:56:38

Input Set : A:\Sequence Listing BioInformatics.txt

Output Set: N:\CRF3\11132000\I698213.raw

```

307 Pro Glu Glu His Pro Val Leu Leu Thr Glu Ala Pro Leu Asn Pro Lys
308          20          25          30
310 Ala Asn Arg Glu Lys Met Thr Gln Ile Met Phe Glu Thr Phe Asn Thr
311          35          40          45
313 Pro Ala Met Tyr Val Ala Ile Gln Ala Val Leu Ser Leu Ala Ser Gly
314          50          55          60
316 Arg Thr Thr Gly Gly Ile Val Leu Asp Ser Gly Asp Gly Val Ser His
317 65          70          75          80
319 Thr Val Pro Ile Tyr Glu Gly Tyr Ala Leu Pro His Ala Ile Leu Arg
320          85          90          95
322 Leu Asp Leu Ala Gly Arg Asp Leu Thr Asp His Leu Met Lys Ile Leu
323          100         105         110
325 Thr Glu Arg Gly Tyr Ser Phe Thr Thr Thr Ala Glu Arg Glu Ile Val
326          115         120         125
328 Arg Asp Met Lys Glu Lys Leu Ser Tyr Ile Ala Leu Asp Phe Glu Gln
329          130         135         140
331 Glu Leu Glu Thr Ser Lys Thr Ser Ser Ser Val Glu Lys Ser Phe Glu
332 145          150         155         160
334 Leu Pro Asp Gly Gln Val Ile Thr Ile Gly Ala Glu Arg Phe Arg Cys
335          165         170         175
337 Pro Glu Val Leu Phe Gln Pro Ser Met Ile Gly Met Glu Asn Pro Gly
338          180         185         190
340 Ile His Glu Thr Thr Tyr Asn Ser Ile Met Lys Cys Asp Val Asp Ile
341          195         200         205
343 Arg Lys Asp Leu Tyr Gly Asn Ile Val Leu Ser Gly Gly Thr Thr Met
344          210         215         220
346 Phe Gly Gly Ile Gly Asp Arg Met Ser Lys Glu Ile Thr Ala Leu Ala
347 225          230         235         240
349 Pro Ser Ser Met Lys Ile Lys Val Val Ala Pro Pro Glu Arg Lys Tyr
350          245         250         255
352 Ser Val Trp Ile Gly Gly Ser Ile Leu Ala Ser Leu Ser Thr Phe Gln
353          260         265         270
355 Gln Met Gln Met Trp Ile Ala Lys Ala Glu Tyr Asp Glu Ser Gly Pro
356          275         280         285
358 Ser Ile Val His Arg Lys Cys Phe
359          290         295

```

VERIFICATION SUMMARY

DATE: 11/13/2000

PATENT APPLICATION: US/09/698,213

TIME: 15:56:39

Input Set : A:\Sequence Listing BioInformatics.txt

Output Set: N:\CRF3\11132000\I698213.raw

L:8 M:270 C: Current Application Number differs, Replaced Current Application No
L:8 M:271 C: Current Filing Date differs, Replaced Current Filing Date
L:94 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:1
L:116 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:2
L:122 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:2
L:128 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:2
L:140 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:2
L:176 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:2
L:179 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:2
L:182 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:2
L:194 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:2
L:213 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:3
L:219 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:3
L:225 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:3
L:237 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:3
L:273 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:3
L:276 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:3
L:279 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:3
L:291 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:3